This is the **revised version** of a draft submitted for possible inclusion in the Proceedings of the Fourth International Conference on the Evolution of Language, Harvard University, March 27-30, 2002

(11,000 words plus 5 figures but including figure captions and 1 table.)

An Action-Oriented Neurolinguistic Framework for the Evolution of Protolanguage

Michael Arbib

Computer Science Department, Neuroscience Program, and USC Brain Project

University of Southern California

Los Angeles, CA 90089-2520

arbib@pollux.usc.edu; http://www-hbp.usc.edu/

Action-Oriented Neurolinguistics in Context

"Action-oriented neurolinguistics" situates the study of brain mechanisms for language within the broader study of neural mechanisms underlying action and perception and the cognitive processes which integrate them. Such an approach raises the question: how much of language rests on mechanisms that evolved specifically to serve language, and how much of language rests on human "inventions" that exploited brain mechanisms that evolved for other purposes? For example, reading involves a set of language-specific mechanisms that are a response to cultural developments of the last few thousand years and must thus exploit brain mechanisms which were extant in early *Homo sapiens* but had evolved for other purposes.

My quest is to understand the brain mechanisms which make it possible for humans to acquire language and to create an evolutionary framework which locates language in relation to the brain mechanisms for action and perception that we share with other primates. This framework is grounded in the Mirror System Hypothesis (explained below) but requires us to go "beyond the mirror" if we are, in future research, to integrate insights from primate studies and human brain imaging to develop an adequate (computational) theory for neurolinguistics. In particular, this requires understanding what is "built in" to the human brain to serve both language acquisition in the child and language performance in the child.

Let us define a *protolanguage* as a system of utterances used by a particular hominid species (possibly including *Homo sapiens*) which we may recognize as a precursor to human language, but which is not itself a human language in the modern sense. (I shall say more of this distinction later.) The approach offered here extends the

Mirror System Hypothesis of Rizzolatti & Arbib (1998), which will be explained below, and provides a neurological basis for the oft-repeated claim that hominids had a (proto)language based primarily on manual gestures before they had a (proto)language based primarily on vocal gestures (e.g., Hewes, 1973; Stokoe, 2001). However, this claim remains controversial and one may contrast two extreme views on the matter:

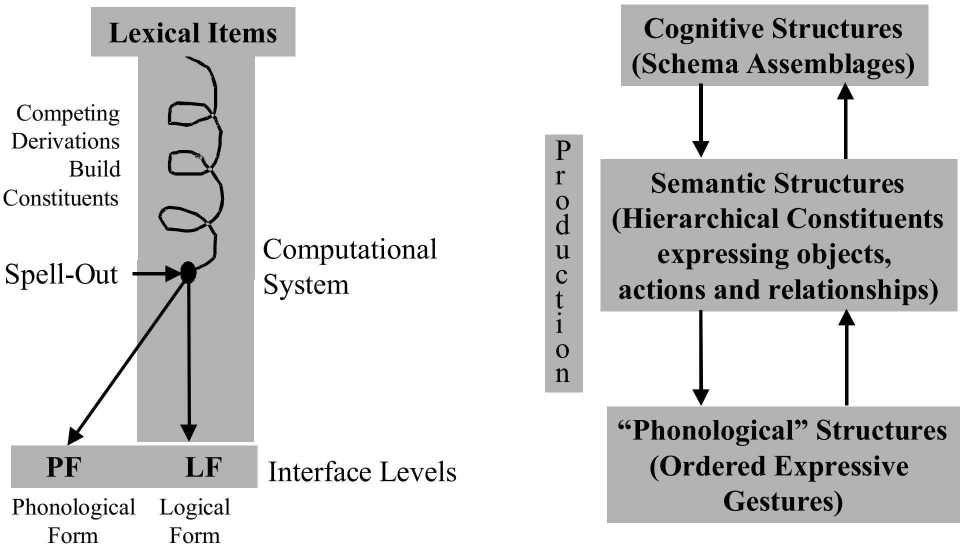
(1) Language evolved directly as speech (MacNeilage, 1998);

(2) Language evolved first as signed language (i.e., as a full language, not protolanguage) and then speech emerged from this basis in manual communication (Stokoe, 2001; Corballis, 2002).

My approach is closer to (2) than to (1). I shall argue that our distant ancestors (e.g., *Homo habilis* through to early *Homo sapiens*) had a protolanguage based extensively on manual gestures ("protosign") which – contra (1) – provided essential scaffolding for the emergence of a protolanguage based primarily on vocal gestures ("protospeech"), but that the hominid line saw advances in both protosign and protospeech feeding off each other in an expanding spiral so that – contra (2) – protosign did not attain the status of a full language prior to the emergence of early forms of protospeech. I will use the term *language-readiness* for *those properties of the brain that provide the capacity to acquire and use language*. Then the hypothesis offered here is that the "language ready brain" of the first *Homo sapiens* supported basic forms of gestural and vocal communication (protosign and protospeech) but not the rich syntax and compositional semantics and accompanying conceptual structures that underlie modern human languages.

Some authors use the term Universal Grammar as a synonym for the ability to acquire language whether or not this rests on innate syntactic mechanisms, but I think we should reserve "grammar" for a mental representation that underlies our ability to combine words to convey and understand meaning, or the systematic description of the commonalities of individual grammars of the members of a community. Some use the term Universal Grammar in the sense of a syntactic framework rich enough to *describe* (most of) the variations seen in all the recorded history of human languages. Yet others go further, and claim that Universal Grammar is a biological property of humans fixed in the ancestral genome of *Homo sapiens*. The most interesting formulation (though I believe it is wrong) of this claim for an innate Universal Grammar is the Principles and Parameters hypothesis, namely that the human genome encodes a grammar, the Universal Grammar, that will yield at least the core syntax of any specific language as soon as certain parameters are set. A further claim here is that these parameters are set so easily on the basis of the child's early experience of language that the Poverty of the Stimulus problem is solved. I espouse an alternative view of what the genome encodes.

Chomsky's Minimalism (Chomsky, 1992), the latest in his series of theories of "competence", characterizes what strings of lexical items are "grammatically correct" as follows (Figure 1a): a set of lexical items is taken at random, the computational system then sees whether legal derivations can be built each of which combines all and only these elements. Spell-Out occurs when one of the legal derivations, if any, is chosen on the basis of some optimality criteria. The Computational System then transforms the result into two different forms, the Phonological Form, the actual sequence of sounds that constitutes the utterance, and the Logical Form, which provides an abstract semantics of the sentence in some quasi-logical notation. There is no attempt here to model actual sentence production – the process starts with words chosen at random and only at the end do we see whether or not they can be arranged in some way that yields a semantic structure.


Lexical Items
Competing Derivations Build Constituents
Spell-Out
Computational System
PF
LF
Interface Levels
Phonological Form
Logical Form
Cognitive Structures (Schema Assemblages)
Production
Semantic Structures (Hierarchical Constituents expressing objects, actions and relationships)
Perception
"Phonological" Structures (Ordered Expressive Gestures)

(a) Chomsky's Minimalism

(b) A Performance Viewpoint

Figure 1. Contrasting (a) Chomsky's Minimalism which provides a "competence characterization" of the phonological and logical forms of well-formed sentences, and (b) a performance viewpoint which relates the phonological and semantic forms of an utterance to cognitive form through production and perception.

Since Chomsky's definition of Grammar has changed so radically over the years, it would seem that whatever was done to argue that an earlier version of Universal Grammar was genetically encoded actually militates against the claim that a Minimalist-style Universal Grammar would be genetically encoded. But might it be countered that the trend of work on Universal Grammar has been such as to make the demands on genetic encoding and the complexity of learning a specific language less demanding with each innovation? To the contrary, Webelhuth (1995, pp.83-85) shows that functional head theory [a key feature of the Minimalist Program] buys increased explanatory adequacy at the price of requiring many additional decisions to be made by language learners – the parameters that distinguish one grammar from another when these grammars are defined by the Minimalist Program are much further from overt language than for the grammars of Chomsky (1965), thus making it less plausible that an innate Universal Grammar solves the Poverty of the Stimulus problem. The Counter Claim here (shared by Deacon, 1997) is that language has evolved to match basic language structures to the learning capabilities of the infant human brain, and that for such a learning system there is indeed a Richness of the Stimulus rather than a Poverty of the Stimulus.

In any case, a neurolinguistic approach should provide a performance approach which explicitly analyzes both perception and production (Figure 1b). Production starts with a Cognitive Form which includes much that we want to talk about; from this is extracted a Semantic Form which structures the objects, actions and relationships to be expressed in the next utterance, and this provides the basis for creating the Phonological Form, the ordered series of expressive gestures (spoken, signed, oro-facial or some combination thereof) which constitutes the overt utterance.[1] Conversely, perception will seek to interpret the phonological form as a semantic form which can then be used to update the perceiver's cognitive form. For example, perception of a visual scene may reveal "Who is doing what and to whom/which" as part of a non-linguistic *action-object frame* in cognitive form. By contrast, the *verb-argument structure* is an overt linguistic representation in semantic form – in modern human languages, generally the action is named by a verb and the objects are named by nouns (or noun phrases). A production grammar for a language is then a specific mechanism (whether explicit or implicit) for converting verb-argument structures into strings of words (and hierarchical compounds of verb-argument structures into complex sentences) and vice versa for perception.

In the brain there may be no single grammar serving both production and perception, but rather a "direct grammar" for production and an "inverse grammar" for perception. Thus the value of a single competence grammar as a reference point for the analysis of perception and production remains debatable. Jackendoff (2002) offers a competence theory with a much closer connection with theories of processing than has been common in generative linguistics and suggests (his Section 9.3) strategies for a two-way dialogue between competence and performance theories. Jackendoff's approach to competence appears to be promising in this regard precisely because it abandons "syntactocentrism", and instead gives attention to the interaction of, e.g., phonological, syntactic and semantic representations.

Language, Protolanguage, and Language-Readiness

Earlier, I defined a *protolanguage* as the system of utterances used by a particular prehominid species (including *Homo sapiens*) which (could we only observe them!) we may recognize as a precursor to human language in the modern sense. Bickerton (1995) has a different definition – for him, a *protolanguage* is a communication system made up of utterances comprising a few words in the current sense placed in a sequence without syntactic structure. Moreover, he asserts that infant language, pidgins, and the "language" taught to apes are all protolanguages in this sense. Bickerton hypothesizes that the protolanguage of *Homo erectus* was also a protolanguage in his sense and that language just "added syntax" through the evolution of Universal Grammar. My counter-proposal is that the "language-readiness" possessed by the first *Homo sapiens* did include the ability to communicate both manually and vocally – I use the terms *protosign* and *protolanguage* for the manual and spoken forms of prelanguage, with the prefix "proto" here having no Bickertonian implication – but that such

[1] Note the formulation "Cognitive Form which includes ..." In most linguistic models of production, it is assumed that a semantic structure is given in some "internal code" and that this must be translated into well-formed sentences. However, in an ongoing conversation, our current mental state and our view of the mental state of our listeners create a richness which our next sentence can only sample, and the generation of that sentence may reflect many factors which change our thoughts even as we express them. To borrow the terminology of motor control, a sentence is not so much a preplanned trajectory as a more or less clumsy attempt to hit a moving and ill-identified target.

protolanguages were composed mainly of “unitary utterances” (a view shared, e.g., by Wray [2002] who relates them to formulaic utterances in modern human languages), and that words co-evolved culturally with syntax through fractionation. The following, very hypothetical, example may clarify what I have in mind (similar examples with much more argumentation are provided by Wray, 2002): Imagine that a tribe has two unitary utterances concerning fire and which, by chance, contain similar substrings which become regularized so that for the first time there is a sign for "fire". Now the two original utterances are modified by replacing the similar substrings by the new regularized substring. Eventually, some tribe members regularize the complementary gestures in the first string to get a sign for "burns"; later, others regularize the complementary gestures in the second string to get a sign for "cooks meat". However, because of the arbitrary origin of the sign for "fire", the placement of the gestures that have come to denote "burns" relative to "fire" differs greatly from those for "cooks meat". It thus requires a further invention to regularize the placement of the gestures in both utterances – and thus as words fractionate from longer strings of gestures, at the same time the protosyntax emerges which combines them.

Criteria for Language-Readiness

I next offer characterizations of language-readiness and language. Of course, both characterizations are preliminary. What is important here is the underlying distinction between what properties the brain has developed as a result of natural, biological selection, and what properties (such as computer programming and space flight) it manifests not because of genetic changes but rather through a long process of enculturation.

We first list the hypothesized properties supporting protolanguage:

LR1: Symbolization: The ability to associate an arbitrary symbol with a class of episodes, objects or actions. (At first, these symbols may not have been words in the modern sense, and they may have been based on manual and facial gestures rather than being vocalized.)

LR2: Intentionality: Communication intended by the utterer to have a particular effect on the recipient.

LR3: Parity (Mirror Property): What counts for the speaker/signer must count (approximately) for the listener/viewer.

The remainder are more general properties, delimiting cognitive capabilities that underlie a number of the ideas which eventually find their expression in language:

LR4: From Hierarchical Structuring to Temporal Ordering: Perceiving that objects and actions have sub-parts; finding the appropriate timing of actions to achieve goals in relation to those hierarchically structured objects.

My point here is that a basic property of language – translating a hierarchical conceptual structure into a temporally ordered structure of words or articulatory gestures (whether signed or vocalized) – is in fact not unique to language but is apparent whenever an animal takes in the nature of a visual scene and produces appropriate behavior.

LR5: Beyond the Here-and-Now 1: The ability to recall past events or imagine future ones.

LR6: Paedomorphy and Sociality: Paedomorphy is the prolonged period of infant dependency which is especially pronounced in humans; this combines with social structures for caregiving to provide the conditions for complex social learning.

Where Deacon (1997) makes symbolization central to his account of the co-evolution of language and the human brain, the present account will stress LR3, since it underlies the sharing of the meaning. I will also argue that only protolanguage co-evolved with the brain, and that the full development of linguistic complexity was a cultural/historical process that required little or no further change from the brains of early *Homo sapiens.* As in this article, Deacon stresses the componential homology (more on this below) which allows us to learn from relations between the brains of monkeys and humans

Criteria for Language

I then suggest that "true language" involves the following further properties:

LA1: Symbolization and Compositionality: The symbols become words in the modern sense, interchangeable and composable in the expression of meaning.

LA2: Syntax, Semantics and Recursion: The matching of syntactic to semantic structures co-evolves with the fractionation of utterances, with the nesting of substructures making some form of recursion inevitable.

Where Hauser, Chomsky & Fitch (2002) asserts that recursion is the one uniquely human component of the faculty of language in the narrow sense(FLN), I stress that recursivity is not restricted to language. Consider the task of chipping one stone with another. The instructions might amount to something like:

(1) Pick a stone to be flaked and a stone to flake it with (the “tool”).

(2) Chip awhile.

(3) Is the flake finished? If so stop.

(4) Is the tool still OK? If so, return to [2]. If not, find a better tool, then return to [2].

It is that notion of “return to some preceding point”, a loop in some abstract structure, that opens the way to recursiveness. In a similar way, once one comes to perceive actions or objects at increasing levels of detail, recursivity follows (recall LR4). The key transition, on this view, is the compositionality that allows cognitive structure to be reflected in symbolic structure (the transition from LR1 to LA1), as when perception (not uniquely human) grounds linguistic description (uniquely human) so that, e.g., the NP [noun phrase] describing a part of an object may optionally form part of the NP describing the overall object. From this point of view, recursion in language is a corollary of the essentially recursive nature of action and perception *once symbolization becomes compositional*.

LA3: Beyond the Here-and-Now 2: Verb tenses or other circumlocutions express the ability to recall past events or imagine future ones.

LA4: Learnability: To qualify as a human language, much of the syntax and semantics of a human language must be learnable by most human children.

I shall return to these criteria in the concluding Discussion

The Mirror System Hypothesis

The brains and bodies of humans, chimps and monkeys differ, and so do their behaviors. They share general physical form and a degree of manual dexterity, but humans have abilities for bipedal locomotion and learnable, flexible vocalization that is not shared by other primates. Moreover, humans can and normally do acquire language, and monkeys and chimps do not – though chimps and bonobos can be trained to acquire a protolanguage (in Bickerton's sense) that approximates the complexity of the protolanguage of a 2 year old human infant. Such "ape protolanguages" are based on hand-eye coordination rather than vocalization. Thus a crucial aspect of human biological evolution has been the emergence of a vocal apparatus and control system that can support speech. But did these mechanisms arise directly from primate vocalizations? The hypothesis presented here is that the route was instead indirect, proceeding via a form of protosign. I will have nothing to say about the biological evolution of the vocal apparatus and will seek instead to delineate some crucial changes in the brain that may have occurred in the course of hominid evolution.

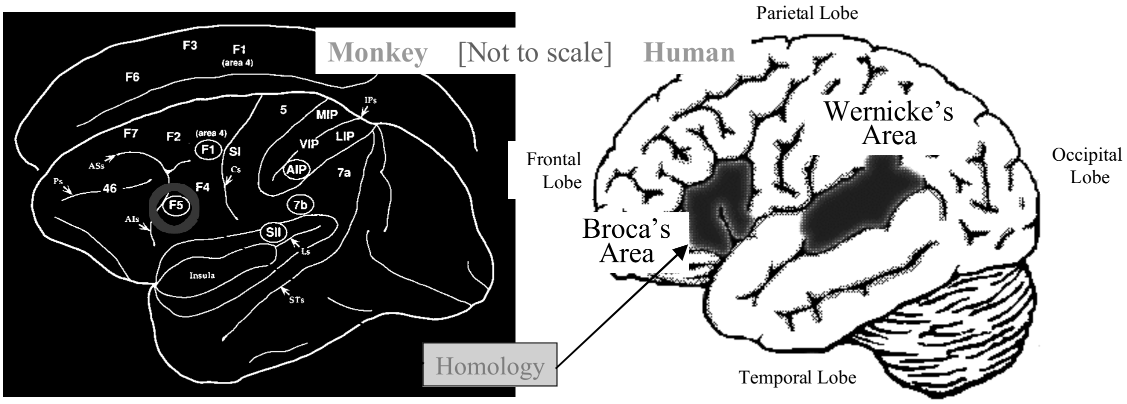


Figure 2. A comparative side view of the monkey brain (left) and human brain (right), not to scale. The view of the monkey brain emphasizes area F5 of the frontal lobe of the monkey; the view of the human brain emphasizes two of the regions of cerebral cortex, Broca's area and Wernicke's area, considered crucial for language processing.

Since humans can learn language and monkeys and chimps cannot, we seek brain regions that are *homologous* (i.e., for there is evidence for a shared ancestry) so that we may learn from analysis of both their similarities and differences. The starting point for the Mirror System Hypothesis is that the system of the monkey brain for visuomotor control of hand movements for grasping has its premotor outpost in an area called F5 (see Figure 2 left) which contains a set of neurons, called *mirror neurons*, such that each mirror neuron is active not only when the monkey executes a specific grasp but also when the monkey observes a human or other monkey execute a more-or-less similar grasp (Rizzolatti et al., 1995). Thus F5 in monkey contains a *mirror system for grasping* which employs a common neural code for *executed* and *observed* manual actions.

The classic papers on the mirror system for grasping in the monkey focus on a repertoire of grasps – such as the precision pinch and power grasp – that seem so basic that it is tempting to think of them as prewired. However, observation of human infants show that little more than the sketchiest coordination of reaching with the location of a visual target plus the palmar grasp reflex is present in the early months of life, and that many months pass before the child has in its motor repertoire the basic grasps (such as the precision pinch) for which mirror neurons have been observed in the monkey. Oztop, Arbib and Bradley (to appear) thus argue that, in monkey as well as human, the basic repertoire of grasps is attained through sensorimotor feedback. They provide the Infant Learning to Grasp Model (ILGM) which explains this process of grasp acquisition. Future modeling will address the issue of how the infant may eventually learn through observation, with mirror neurons and grasping circuitry developing in a synergistic manner (see Zukow-Goldring, Arbib and Oztop, to appear, for a study of infant skill acquisition and further discussion).

The next few paragraphs are designed to show that the functionality of mirror neurons rests on the embedding of F5 in a much larger neural system, and briefly outlines the Mirror Neuron System One model (MNS1, Oztop and Arbib, 2002) which demonstrates how neural plasticity can yield mirror neuron functionality through correlated experience rather than through "pre-wiring". Readers who wish to omit this part of the argument may go directly to the paragraph starting "Oztop and Arbib (2002) provide …"

To introduce this model, it is first useful to distinguish the *mirror neurons* in area F5 of monkey premotor cortex, active both when the monkey executes a specific grasp and when it observes an other executing a more-or-less similar grasp, from *canonical neurons* in F5 which are active when the monkey itself is doing the grasping in response to sight of an object but not when the monkey sees someone else do the grasping. More subtly, canonical neurons fire when they are presented with a graspable object, irrespective of whether the monkey performs the grasp or not – but clearly this must depend on the extra condition that the monkey not only sees the object but is "aware", in some sense that it is possible to grasp it. Were it not for this caveat, canonical neurons would also fire when the monkey observed the object being grasped by another.

Taira et al. (1990) established that AIP (Figure 2 left; called AIP because it is in the Anterior part of the Intraparietal sulcus of monkey Parietal cortex[2]) extract neural codes for "affordances" for grasping from the visual stream and sends these on to area F5. *Affordances* (Gibson, 1979) are features of the object relevant to action, in this case to grasping, rather than aspects of identifying the object's identity. For example, a screwdriver may be grasped by the handle or by the shaft, but one does not have to recognize that it is a screwdriver to recognize its different affordances. These affordances provide the input to the canonical neurons of F5:

(1) $\text{AIP} \rightarrow \text{F5}_{\text{canonical}}$

The FARS model (Fagg and Arbib, 1998) provides a computational account of the neural mechanisms, including those in (1), for going from the shape of part of an object to the grasping of that part of the object. By contrast, the task for the mirror system is to determine whether the shape of a hand and its trajectory are "on track" to grasp an observed affordance of an object, and so we have to find other regions of the brain that provide appropriate visual input. One relevant brain region in the parietal cortex is PF which contains neurons responding to the sight of goal

[2] The reader unfamiliar with the neuroanatomy of the monkey brain need not worry about the details of cerebral localization or the anatomical labels in Figures 2 and 3. The only point important here is that parietal cortex (the "top right" part of both monkey and human cerebral cortex in Figure 2) can be subdivided into many regions and that different parietal regions provide the input to the canonical and mirror neurons, one (AIP) concerned with the affordances of objects and the other (PF) concerned with relations between an object and the hand that is about to grasp it.

directed hand/arm actions (Fogassi et al., 1998). MNS1 (Figure 3) is organized around the idea that (1) is complemented by

(2) PF → $F5_{mirror}$

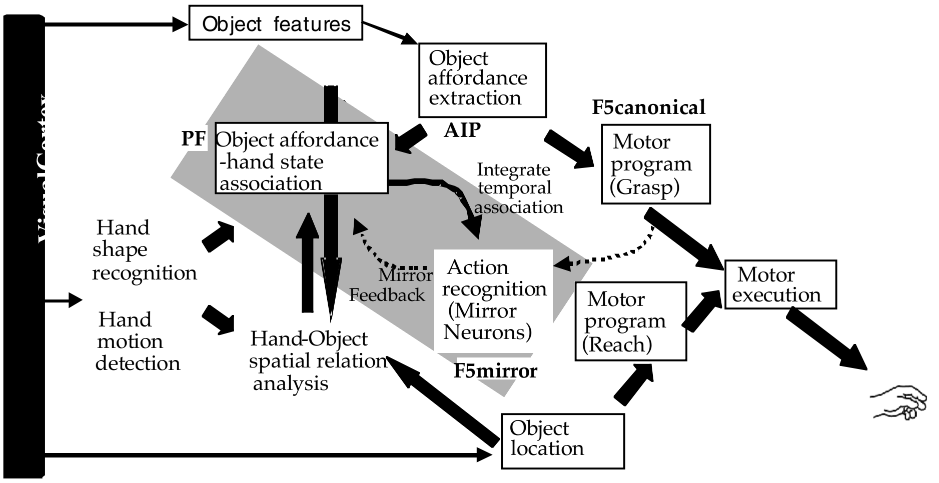

Object features
Object affordance extraction
F5canonical
PF
Object affordance -hand state association
AIP
Motor program (Grasp)
Integrate temporal association
Hand shape recognition
Mirror Feedback
Action recognition (Mirror Neurons)
Motor execution
Hand motion detection
Hand-Object spatial relation analysis
Motor program (Reach)
F5mirror
Object location

Figure 3. A schematic view of the Mirror Neuron System. The Mirror Neuron System One (MNS1) model (Oztop and Arbib, 2002) focuses on the circuitry highlighted by the grey diagonal rectangle. (i) Top diagonal: Object features are processed by AIP to extract grasp affordances, these are sent on to the canonical neurons of F5 that choose a particular grasp. (ii) Bottom right. Recognizing the location of the object provides parameters to the area which programs arm movements during the reach. The information about the reach and the grasp is taken by motor cortex to control the hand and the arm. (iii) Essential elements for the mirror system: Bottom left are two schemas, one to recognize the shape of the hand of the actor being observed by the monkey whose brain we are interested in, and the other to recognize how that hand is moving. Just to the right of these is the schema for hand-object spatial relation analysis. It takes information about object features, the motion of the hand and the location of the object to infer the relation between hand and object. Just above this is the schema for associating object affordances and hand state. Together with F5 canonical neurons, this last schema (posited to be in PF) provides the input to the F5 mirror neurons.

As shown in Figure 3 (see the caption for details), the MNS1 model shows how the interaction of various brain regions provide mechanisms to evaluate the key criteria for activating a mirror neuron:

- The preshape that the monkey is seeing corresponds to the grasp that the mirror neuron encodes.
- The preshape that the observed hand is executing is indeed appropriate to an affordance of the object that the monkey can see (or remember).
- The hand must be moving on a trajectory that will indeed bring it to grasp the part of the object that provides that affordance.

Oztop and Arbib (2002) provide an explicit computational model of how the mirror system may learn to recognize the hand-object relations associated with grasps already in its repertoire. The details are irrelevant here. What is relevant is that such learning models, and the data they address, make clear that *mirror neurons are not restricted to recognition of an innate set of actions but can be recruited to recognize and encode an expanding repertoire of novel actions.*

With this background on the monkey, I now turn to the question “Is there a mirror system for grasping in the human brain?” Brain imaging cannot answer the question of whether the human brain contains neurons each matched to specific grasps. What it can do is seek brain regions that are more active both for grasping an object and for observation of grasping an object when compared to simple observation of an object. Dramatically, the prefrontal area with this "mirror property" in humans is Broca's area (see Figure 2 right), the frontal area most strongly implicated in the production of language. Moreover, there is evidence that part of Broca’s area is homologous to monkey F5 (Rizzolatti and Arbib,1998, [henceforth R&A] to whom the reader is referred for references to the primary data). This led R&A to formulate:

The Mirror System Hypothesis: The parity requirement for language in humans – that what counts for the speaker must (approximately) count for the hearer – is met because language evolved from the mirror system for grasping in the common ancestor of monkey and human with its capacity to generate and recognize a set of manual actions.

As such, the Mirror System Hypothesis provides a neural "missing link" to those (Stokoe, 2001, for a recent example) who, struck by the ability of deaf children to acquire sign language as readily as normal children acquire speech, have argued not only that language must be considered as multi-modal – with manual gesture on a par with speech – but also that some form of signing could have provided the scaffolding for the evolution of speech. They stress that pantomime can convey a wide range of meanings without recourse to the arbitrary associations that underlie the sound-meaning pairings of most words of a spoken language.

It could be tempting to hypothesize that certain species-specific vocalizations of monkeys (such as the snake and leopard calls of vervet monkeys) provided the basis for the evolution of human speech. However, these primate vocalizations appear to be related to the cingulate cortex[3] rather than the F5 homologue of Broca's area. I think it likely (though empirical data are sadly lacking) that the primate cortex contains a mirror system for such species-

[3] Adequate exposition of such terms as cingulate cortex would unduly burden the present argument. What matters here is that we might expect language to have evolved “in the speech domain” and that if so we might expect Broca’s area to be homologous to the monkey’s “vocalization cortex” – but it is not, and that drives the Mirror System Hypothesis.

specific vocalizations, and that a related mirror system persists in humans, but I suggest that it is a complement to, rather than an integral part of, the speech system that includes Broca's area in humans.

The Mirror System Hypothesis is a neurolinguistic hypothesis – it is an account of the evolution of the brain mechanisms that give humans a language-ready brain. It claims that a *specific* mirror system – the primate mirror system for grasping – evolved into a key component of the mechanisms that render the human brain language-ready. It is this specificity that will allow us to explain below why language is multi-modal, its evolution being based on the execution and observation of hand movements. To achieve these implications we must go beyond the core data on the mirror system to stress that manipulation inherently involves hierarchical motor structures which are unavailable for the closed call system of primates. The mastery of hierarchical motor structures is not the property of F5 alone but involves its integration within a network of distributed cortical and subcortical structures. To understand this point, we recall Lashley's (1951) critique of behaviorist psychology: If we tried to learn a sequence like A → B → A → C by reflex chaining, what is to stop A triggering B every time, to yield the performance A → B → A → B → A → …? A solution is to store the "action codes" (motor schemas) A, B, C, … in one part of the brain and have another area hold "abstract sequences" and learn to pair the right action with each element, as in Figure 4.[4]

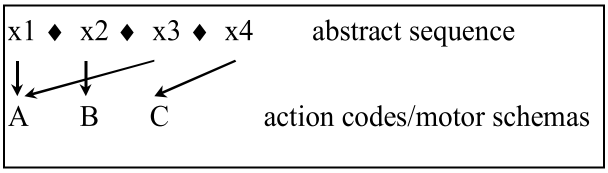


Figure 4. A solution to the problem of serial order in behavior: Store the "action codes" A, B, C, … in one part of the brain and have another area hold "abstract sequences" and learn to pair the right action with each element:

Having established what the Mirror System Hypothesis does say, it is equally important to emphasize what the Mirror System Hypothesis does not say:

(i) It does not say that having a mirror system is equivalent to having language. Monkeys have mirror systems but do not have language. Moreover, we expect that many species have mirror systems for varied socially relevant behaviors (cf. birdsong, primate calls, etc.)

[4] For those readers conversant with neurophysiology, I add that in my group's modeling of visuomotor control of grasping (Fagg and Arbib, 1998) we posit that the action codes are in F5, the sequence information is in the region of the supplementary motor area called SMA-proper, and that the management of these sequences involves essential activity of the basal ganglia to manage priming and inhibition. The important point for all readers is that no single brain region holds the "magic key" for perception and action, let alone language. Both F5 in monkey and Broca's area in human are embedded in much larger neural systems. Indeed, Lieberman (2002) downplays the role of Broca's area in language and emphasizes the role of the basal ganglia (the brain region from which different defects may yield Parkinson's disease and Huntington's disease).

(ii) It does not say that the ability to match the perception and production of single gestures is sufficient for language.

(iii) It does not say that language evolution can be studied in isolation from cognitive evolution more generally. In using language, we make use of, for example, negation, counterfactuals, and (in some languages) verb tenses. But these linguistic structures are of no value unless we can understand that the facts contradict an utterance, and can recall past events and imagine future possibilities.

From Grasp to Language: Seven Hypothesized Stages of Evolution

How, then, do we get from the brain of our common ancestor with monkeys via the brain of our common ancestor with chimpanzees to the language-ready brain of the first *Homo sapiens*? Arbib (2002) extends the Mirror System Hypothesis to present seven stages of this evolution, with imitation grounding two of the stages. The first three stages are pre-hominid:

S1: Grasping;

S**2:** A mirror system for grasping shared with the common ancestor of human and monkey; and

S3: A simple imitation system for grasping shared with common ancestor of human and chimpanzee.

The next three stages then distinguish the hominid line from that of the great apes:

S4: A complex imitation system for grasping,

S5: *Protosign*, a manual-based communication system, breaking through the fixed repertoire of primate vocalizations to yield an open repertoire

S6: *Proto-speech*, resulting from the ability of control mechanisms evolved for protosign coming to control the vocal apparatus with increasing flexibility.

The final stage is claimed to involve little if any biological evolution, but instead to result from cultural evolution (historical change) in *Homo sapiens*:

S7: *Language*: the change from action-object frames to verb-argument structures to syntax and semantics; the co-evolution of cognitive and linguistic complexity.

Arbib (2002) discusses the distinction between "simple" and "complex" imitation – with chimpanzees being capable of the former but not the latter. Rather than rehears these details here, it is important to distinguish two roles for imitation in the transition from stage **S4** to stage **S5**:

R&A stress the transition from *praxic action* directed towards a goal object to *pantomime* in which similar actions are produced away from the goal object. But communication is about far more than grasping! To pantomime the flight of a bird you might move your hand up and down in a way that indicates the flapping of a wing. To do this, you must move beyond mere imitation of hand movements – your pantomime uses movements of the hand (and arm and body) to imitate movement other than hand movements. You can also pantomime an object by movements which suggest tracing out the characteristic shape of the object. Imitation is the generic attempt to reproduce movements performed by another, whether to master a skill or simply as part of a social interaction. By contrast, pantomime is performed with the intention of getting the observer to think of a specific action or event. It is essentially communicative in its nature. The imitator observes; the panto-mimic intends to be observed. This is where the intentionality (LR2) of language-readiness comes in.

The transition to pantomime does seem to involve a genuine neurological change. Mirror neurons for grasping in the monkey will fire only if the monkey sees *both* the hand movement and the object to which it is directed (Umilta et al., 2001). A grasping movement that is not made in the presence of a suitable object, or is not directed toward that object, will not elicit mirror neuron firing. By contrast, in pantomime, the observer *infers* the goal or object of the action from observation of the movement in isolation.

A further critical (but perhaps non-biological) change en route to language emerges from the fact that in pantomime it might be hard to distinguish a movement signifying “bird” from one meaning “flying". This would favor the invention of abstract gestures available as elements for the formation of compounds which can be paired with meanings in more or less arbitrary fashion. This requires extending the mirror system to attend to a whole new class of hand movements, those with conventional meanings agreed upon by the protosign community to reduce ambiguity and extend semantic range. (As ahistorical support for this, one might note the way that certain noun/verb pairs are differentiated by movement in American Sign Language, ASL. For example, Supalla and Newport, 1978, note that AIRPLANE is signed in ASL with tiny repeated movements of a specific handshape, while FLY is signed by moving the same handshape along an extended trajectory. [I say “ahistorical” above because such signs are part of a modern human language rather than holdovers from protosign. Nonetheless, they exemplify the mixture of iconicity and convention that, I claim, distinguishes protosign from pantomime.])

Kohler et al. (2002) have studied mirror neurons for actions which are accompanied by characteristic sounds, and found that a subset of these are activated by the sound of the action (e.g., breaking a peanut in half) as well as sight of the action. Does this suggest that protospeech mediated by the F5 homologue in the hominid brain could have evolved without the scaffolding provided by protolanguage? My answer is negative for three reasons. First, I have argued that imitation is crucial to grounding pantomime in which a movement is performed in the absence of the object for which such a movement would constitute part of a praxic action. However, the sounds studied by Kohler et al. (2002) cannot be created in the absence of the object and there is no evidence that monkeys can use their vocal apparatus to mimic the sounds they have heard. Second, I have noted (and see Stokoe 2001 for further argument) that pantomime allows a broad range of communication which does not require an agreed on convention between sender and receiver. However, I have further stressed that such pantomime is limited, and that the range of communication can be greatly improved by the development of conventions on the use of gestures that do not directly pantomime anything, but instead are developed by a community to refine and annotate the more obvious forms of pantomime – forms which themselves would become increasingly ritualized with use. The notion, then, is that the manual domain supports the expression of meaning by sequences and interweavings of gestures, with a progression from “natural” to increasingly conventionalized gesture to speed and extend the range of communication within a community.

I would then argue that Stage 5 (protosign) provides the scaffolding for Stage 6 (protospeech). We have already seen that some mirror neurons in the monkey are responsive to auditory input. We now note that there are orofacial neurons in F5 that control movements that could well affect sounds emitted by the monkey. The speculation here is that the evolution of a system for voluntary control of intentional communication based on F5/Broca's area could then lay the basis for the evolution of creatures with more and more prominent connections from F5/Broca’s area to the vocal apparatus. This in turn could provide conditions that lead to a period of co-evolution of the vocal apparatus and the neural circuitry to control it. Corballis (2002) offers cogent reasons for the selective advantage of incorporating vocalization in an originally hand-based communicative repertoire and I will not attempt to summarize them here.

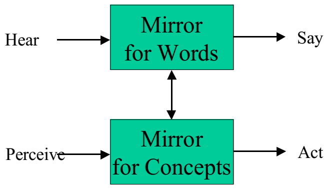


Figure 5. The Saussurean sign - linking word and meaning (adapted from Hurford, 2003).

Figure 5 is based on a scheme offered by Hurford (2003) in his critique of the Mirror System Hypothesis. I shall argue that the top row of the figure is indeed supported by the Mirror System Hypothesis, but that the bottom row is not. For the top row, consider how the Mirror System Hypothesis supports the transition from a mirror system for grasping in F5 in the common ancestor of monkey and human to a mirror system for words (in the sense of composites of articulatory gestures) in Broca’s area in humans:

- Mirror system for grasping and manual praxic actions.
- Mirror system for pantomime of grasping and manual praxic actions.
- Mirror system for pantomime of actions outside the panto-mimic's own behavioral repertoire (e.g., flapping the arms to mime a flying bird).
- Mirror system for conventional gestures used to formalize and disambiguate pantomime (e.g., to distinguish "the bird" from "the flying")
- Mirror system for all manual (and related facial) communicative gestures
- Mirror system for all vocal (and related manual and facial) communicative gestures.

Hurford (2003) argues that “if humans are organized in this respect like macaques, the mental representation of the concept GRASP/GRASPING involves some neurons which are involved both in the act of grasping and in the observation of grasping. So thinking of grasping (either by oneself or by someone else) activates these mirror neurons. Similarly, it seems likely that a representation of the concept WALK/WALKING will involve mirror neurons involved both in the observation and the performance of walking." To extend this to objects, Hurford suggests that "representations of objects involve some congruence between motor and sensory neurons, similar to that found in the representations of actions. ... The mental representations of tools involve areas of motor cortex appropriate for handling them, beside sensory information about what the tools look like[.] … Similarly, one's concept of, say, an apple, includes motor information about how to hold it and bite it, as well as sensory information about what it looks/tastes/smells like." The suggestion here seems to be that there is a mirror system for all concepts – actions, objects and more besides – which links the perception and action related to each concept. Contrary to this, I argue that the bottom row of Figure 5 takes us into territory on which the Mirror System Hypothesis has remained silent. First, the Mirror System Hypothesis is agnostic as to whether there are mirror systems for actions other than those bulleted above and, if so, where these might be located in the human brain. Moreover, I do not believe that there is a mirror system for all concepts. In schema theory (Arbib, 1981; 2003a), I distinguish between *perceptual schemas* which determine whether a given "domain of interaction" is present in the environment and provide parameters concerning the current relationship of the organism with that domain, and *motor schemas* which provide the control systems which can be coordinated to effect a wide variety of actions. Schema instances may be combined (possibly with those of more abstract schemas, including coordinating schemas) to form new schemas as "schema assemblages". To see why perceptual and motor schemas are in general kept separate in this account, consider that recognizing an object (an apple, say) may be linked to many different courses of action (to place the apple in one's shopping basket; to place the apple in the bowl at home; to peel the apple; to eat the apple; to discard a rotten apple, etc.). Of course, once one has decided on a particular course of action then specific perceptual and motor subschemas may be invoked. But note that, in the list just given, some items are apple-specific whereas other invoke generic schemas for reaching and grasping. It was considerations like this that led me to separate perceptual and motor schemas – a given action may be invoked in a wide variety of circumstances; a given perception may, as part of a larger assemblage, precede many courses of action. Putting it another way, there is no one "grand apple schema" which links all "apple perception strategies" to "every act that involves an apple".

Thus I reject the notion of a mirror system for concepts. Instead, I visualize the brain as encoding a varied network of perceptual and motor schemas. Only rarely (as in the case of certain basic actions) will the perceptual and motor schemas be integrated into a "mirror schema". In general, a word may be linked to many schemas, with varying context-dependent activation strengths. On this view, I do not see a "concept" as corresponding to one word, but rather to a graded set of activations of the schema network.

With this we can return to Figure 5. Hurford (2003) asks what, in neural terms, the bidirectional arrow might be, stressing that the "arbitrariness of the sign" implies that in general there is no overlap between the neurons involved in the representation of the meaning and those involved in the representation of the sound. He thus concludes that mirror neurons do not help to account for the facility shown by humans in acquiring a large vocabulary. Indeed, they do not *of themselves*, which is why I introduced the two stages of imitation leading to pantomime and the conventionalization of gesture as the route from merely having a mirror system to having a brain that could support protosign. The bullets above trace how, throughout this process of evolution of multiple interacting brain regions, a mirror system for grasping became transformed to provide a mirror system for the production and perception of symbols while retaining its ancestral role in grasping *per se*. Stage S6 of the Mirror System Hypothesis explains why, counterintuitively, protosign may have provided essential scaffolding for the emergence of protospeech and the language-ready brain. The flexibility of the resulting system is provided by the observation (supported by the earlier discussion of the MNS1 model) that the mirror system repertoire is not pre-wired but instead emerges through experience.

A major goal for our action-oriented neurolinguistics is then to understand what additional mechanisms are employed by the human brain in linking Cognitive Form to Semantic Form and Phonological Form, in part by responding to the challenges in extending the model of Figure 4. Returning to Figure 5, we also need to explain (i) why it is easy for humans to build a "mirror system" for "words", and (ii) how this mirror system can be linked to the perceptual and motor schemas for concepts. In each case, I suggest that this is not a general property spread across the human brain, but instead involves different patterns of plasticity linked to specific brain mechanisms which evolved along the hominid line.

With this, I complete my presentation of the six stages of evolution of the language-ready brain. The details of the passage from protolanguage to language, the posited post-biological stage S7 above, are beyond the scope of this article. A preliminary account of the possible transitions from action-object frame to verb-argument structure to syntax and semantics is given in the Arbib (2002). A far more thorough account is in preparation.

Discussion

To conclude, we revisit the ten properties, LR1-LR6 and LA1-LA4, hypothesized to support protolanguage and "true language" and try to reconcile them with the seven stages, S1-S7 of the extended Mirror System Hypothesis. Table 1 re-orders these elements in a fashion designed to facilitate this comparison. The rows with the right hand cell filled show the way in which the Mirror System Hypothesis provides an action-oriented framework for the evolution of protolanguage.

Table 1

A comparative view of the six properties, LR1-LR6, of protolanguage and the 4 further properties, LA1-LA4, of language (left column) and the seven stages, S1-S7, of the extended Mirror System Hypothesis (right column).

LR4: From hierarchical structuring to temporal ordering	**S1: Grasping** **S2: Mirror system for grasping**
	S3: Simple imitation **S4: Complex imitation**
LR1: Symbolization **LR2: Intentionality** **LR3: Parity (Mirror Property)**	**S5: Protosign**
	S6: Proto-speech
LA1: Symbolization and compositionality **LA2: Syntax, semantics and recursion**	**S7: Language**
LR5: Beyond the here-and-now 1 **LA3: Beyond the here-and-now 2**	
LR6: Paedomorphy and sociality **LA4: Learnability**	

LR4//S1;S2: Where Hauser et al. (2002) view recursion as the uniquely human key to the faculty of language, LR4 (From Hierarchical Structuring to Temporal Ordering) reminds us that the study of animal behavior is replete with examples of how an animal can analyze a complex sensory scene and, in relation to its internal state, determine a course of action. When a frog faced with prey and predators and a barrier ends up, say, choosing a path around the barrier to escape the predator (see Cobas and Arbib, 1992, for a model) it exemplifies the ability to analyze the spatial relation of objects in planning its action. However, there is little evidence of recursion here – once the frog rounds the barrier, it seems to need to see the prey anew to trigger its prey-catching behavior. By contrast, the flow diagram given by Byrne (2003) shows that the processing (from getting a nettle plant to putting a folded handful of leaves into the mouth) used by a mountain gorilla when preparing bundles of nettle leaves to eat is clearly recursive. Gorillas (like many other species, and not only mammals) have the working memory to refer their next action not only to sensory data but also to the state of execution of some current plan. Thus when we refer to the monkey's grasping and ability to recognize similar grasps in others (S1 and S2) it is a mistake to treat the individual grasps in isolation – the F5 system is part of a larger system that can direct those grasps as part of a larger recursively structured plan. Of course, more needs to be done in understanding the interaction of the diverse neural systems that support grasping and how they each change in the evolution of the language-ready brain.

//S3;S4: The multi-causal view of evolution implicit in the above argument is (to simplify) that a change in one part of a complex system (e.g., in working memory) can be exapted to yield improved performance of another system (e.g., planning) in a way which makes a mutation affecting another system (e.g., perception of patterns of social behavior) beneficial even though it would not have adaptive value had not prior changes (some primary and some secondary) already occurred. Such changes continue in a cascade over tens of millennia. What offers promise of an account in this vein becoming more than a series of just-so stories (which have value in framing our arguments) is the method of comparative biology – we cannot study the brains of our ancestors, but we can compare our brains with those of other species in the search for meaningful homologies (Arbib and Bota, 2003) which can anchor and be anchored by the inference of evolutionary relationships. Similarly, we can compare behaviors. The fact that monkeys have little or no capacity for imitation and had their common ancestor with humans some 20 million years ago, while chimpanzees have an ability for "simple" imitation and had their common ancestor with humans some 5 million years ago makes plausible that our evolutionary path took us through the emergence of simple imitation (S3) before 5 million years ago, and the emergence of complex evolution (S4) more recently. The left-hand cell is blank in this row of the table because one can certainly entertain many scenarios for why both of these stages would have been adaptive in ways that had no relation to language as such. However, these stages are crucial to the present theory because complex imitation is central to the ability of human infants to acquire language and behavior in a way that couples these in their increasing mastery of the physical and social world.

LR1;LR2;LR3//S5: The Mirror System Hypothesis as presented above assumes, rather than provides an explanation, for LR2, the transition from making praxic movement, e.g., those involved in the immediate satisfaction of some appetitive or aversive goal and those intended by the utterer to have a particular effect on the recipient. As a placeholder, let me note that (as Darwin [1872/1965] observed long ago) the facial expressions of conspecifics provide valuable cues to their likely reaction to certain courses of behavior (a rich complex summarized as "emotional state") and that this ability can be observed across a far wider range of mammalian species than just the primates. Moreover, the F5 region contains orofacial cells as well as manual cells. One might thus posit a progression from control of emotional expression by systems that exclude F5 to the extension of F5's mirror capacity from manual to orofacial movement and then, via its posited capacity (achieved by stage S3) for simple imitation, to support the imitation of emotional expressions. This would then provide the ability to affect the behavior of others by, e.g., appearing angry. This would in turn provide the evolutionary opportunity to generalize the ability of F5 activity to affect the behavior of conspecifics from vocal expressions to a general ability to use the imitation of behavior (as distinct from praxic behavior itself) as a means to influence others. This in turn makes possible reciprocity by a process of backward chaining where the influence is not so much on the praxis of the other as on the exchange of information. With this, the transition described by LR2 (intentionality) has been achieved in tandem with the achievement of LR1, the ability to associate an arbitrary symbol with a class of episodes, objects or actions (but without, at this stage, compositionality). In the present theory, the crucial ingredient in LR1 is the extension of imitation from the imitation of hand movements to the ability to in some sense project the degrees of freedom of movements involving other effectors (and even non-humans or, say, the passage of the wind through the trees) to create hand movements that could evoke something of the original in the brain of the observer. As our discussion of Hurford showed, this involves not merely changes internal to the mirror system but its integration with a wide range of brain regions involved in the elaboration and linkage of perceptual and motor schemas. With this, LR3 (parity; the mirror property) follows automatically – what counts for the signer must count (approximately) for the observer.

The transition to protosign (S5) may not require further biological changes but does involve the discovery that it is more efficient to use conventionalized gestures for familiar objects, actions and episodes than to initiate an original pantomime. With time and within a community these gestures would become increasingly stylized and their link to the original pantomime would be lost. But this loss would be balanced by the discovery that when an important distinction cannot be conveniently pantomimed an arbitrary gesture may be invented to express the distinction. Deixis presumably plays a crucial role here – what cannot be pantomimed may be shown when it is present so that the associated symbol may be of use when it is absent. Protosign, then, emerges as a manual-based communication system rooted originally in pantomime but which is open to the addition of novel communicative gestures as the life of the community comes to define the underlying concepts and makes it important to communicate about them.

//S6: I have placed S6, the evolution of protospeech, in a separate row from S5, the evolution of protosign, to stress the point that the role of F5 in grounding the evolution of a protolanguage system would work just as well if we and all our ancestors had been deaf. However, primates do have a rich auditory system which contributes to species survival in many ways of which communication is just one (Ghazanfar, 2003). The hypothesis here, then, is not that the protolanguage system had to create the appropriate auditory and vocal-motor system "from scratch" but rather that it could build upon the existing mechanisms to derive protospeech. My hypothesis is that protosign grounded the crucial innovation of using arbitrary gestures to convey novel meanings, and that this in turn provided the scaffolding for protospeech. Consistent with my view that true language emerged during the history of *Homo sapiens* and the observation that the vocal apparatus of humans is especially well adapted for speech, I suggest that the interplay between protospeech and protolanguage was an expanding spiral which yielded a brain that was ready for language in the multiple modalities of gesture, vocalization, and facial expression.

LA1;LA2//S7: I claim that stage S7, the transition from protolanguage to language, is the culmination of manifold discoveries in the history of mankind, arguing that it required manifold incremental changes to yield the full structure of language: The symbols developed so painfully in prehuman societies become words in the modern sense, interchangeable and composable in the expression of meaning (LA1). This fractionation of the vocabulary made necessary the development of syntax and semantics to gain the benefits of putting the pieces together in novel combinations without an explosion of ambiguity (LA2). What I stress is that there was no sudden transition from holophrastic utterances to an elaborate system of principles and parameters. Rather, languages emerged through a process of bricolage (tinkering) which yielded many novelties to handle special problems of communication, with a variety of generalizations amplifying the power of groups of inventions by unifying them to provide tools of greatly extended range.

I recently had the opportunity to review (Arbib, 2003b) the collection of papers on *Linguistic Evolution through Language Acquisition: Formal and Computational Models* edited by Ted Briscoe (2002). Some remarks based on that review may reinforce the point I am trying to make here. This book comprises two very different sets of chapters. Those by Worden, Batali, Kirby and Hurford (WBKH) use computer simulation to demonstrate that agents with no innate syntactic structure can interact to create and then extend both the lexicon and syntax of languages over many generations. On the other hand, those by Niyogi, Turkel and Briscoe (NTB) all start with an innate Universal Grammar characterized by a set of principles and parameters. The job of the language learner is to infer the parameter settings of the strings it encounters, or to reach consensus within a population of agents as to which parameter settings to converge upon in resolving initially disparate data. I have already explained above why I reject this approach – further reasons are given in my review. Here let me instead focus on WBKH who all assume that the agents under study have a great deal of “innate” language-related structure. The “linguistic evolution” studied by WBKH is not the evolution of the *capacity* for language but rather the ability of a community to build on innate mechanisms to seek coherence in the encoding of meanings by strings of symbols. The “through language acquisition” of the book’s title is that as each agent is added to a population it seeks to model its sample of the utterances of the existing population and in so doing changes the population profile. The language defined by the statistics of the population’s output “evolves” accordingly.

The “language-readiness” assumed by WBKH is that each agent starts with the same innate sets M of structured meanings (e.g., semantic trees or logical expressions over some small set of symbols) and S of strings of symbols and that they are equipped with learning algorithms whose sole purpose is to infer rules that can generate meaning to string encodings that increasingly match the encodings used by other individuals. Further, WBKH assume that the learner has direct access to the meaning of each string, i.e., its training data take the form of (meaning, string) pairs. (They assume stronger innate mechanisms than I feel are warranted in the human case, but the key point here is that they assume nothing like a Universal Grammar.) A population creates and learns (meaning, string) pairs. At first, the assignment of string to meaning is essentially random and varies from individual to individual, but as each new member is added to a population it will seek to learn pairings that are already in use. WBHK show that as "old" agents leave the population and "new" agents join it, something like a coherent language emerges across the course of generations. The resulting consensus assignment gains its power by reflecting semantic structure in the structure of the corresponding string. A coherent pattern emerges that has never before existed in the history of the "species". Compositionality and recursion in the mapping emerge across generations as a result of general conditions of learnability and coherence, rather than being built in as innate principles. The various WBHK mechanisms are indeed supportive of S7 though we also need "post-biological" modeling to show how language might guide the discovery of concepts, rather than expressing concepts that are already built-in, let alone formalized. Another crucial target is the unification of such studies with actual data on the linguistic and cognitive development of children. The WBKH methodology is "Here is a set of interacting agents. Here are some measures of how the 'languages' of the agents in a population do or do not cohere with one another. Here are simulation results showing interesting patterns of increasing coherence across the generations" The patterns are indeed interesting but there are no studies of the form "Here is a model of language acquisition. Here are real data on language acquisition in human children. Look how well the model explains key aspects of the data. It is plausible that these mechanisms have long been present in hominids and so could account for historical patterns of the emergence of language as well as its ontogeny in the individual child."

LR5;LA3//: LR5 and LA3 together make two different points: The first point is that language involves many powerful devices that extend the range of communication but that might not be considered as touchstones to the definition of language. Thus if one took a human language and removed all reference to time one might still want to call it a language rather than a protolanguage, even though one would agree that it was thereby greatly impoverished. Similarly, the number system of a language can be seen as a useful, but not definitive, "plug in". LA3 nonetheless suggests that the ability to talk about past and future is a central part of human languages as we understand them. Secondly, LR5 reminds us that these features of language would be meaningless (literally) without the underlying cognitive machinery – in this case the substrate for episodic memory provided by the hippocampus (Burgess, Jeffery, and O'Keefe, 1999) and the substrate for planning provided by frontal cortex (Passingham, 1993, Chapter 10). Thus the neurolinguist must not only seek to learn from the syntactician how time is expressed in a variety of languages but also seek to understand how these verbal structures are linked to the cognitive structures which give them meaning and thus, presumably, grounded their evolution – irrespective of what autonomy the syntactic structures may have when severed from the contingencies of communication.

LR6;LA4//: The final row of the table again links a biological condition with a "supplementary" property of human languages. This supplementary property is that languages do not simply exist – they are acquired anew (and, as we saw above, may be slightly modified thereby) in each generation (LA4). The biological property is an inherently social one about the nature of the relationship between parent (or other caregiver) and child (LR6) – the prolonged period of infant dependency which is especially pronounced in humans has co-evolved with the social structures for caregiving that provide the conditions for the complex social learning that makes possible the richness of human cultures in general and of human languages in particular.

I deny that there is any single “magic mutation” or change of brain or body structure that created the capacity for language as humans know it today, and the rows in Table I with the right hand cell empty indicate some of the complementary work that must be done. However, the above discussion should make clear that the evolution of the mirror system is one important aspect of brain changes underlying the evolution of language-readiness and that the Mirror System Hypothesis does indeed provide an action-oriented framework for the evolution of protolanguage which should in future serve to anchor new contributions to linguistics through enriched attention to neurolinguistics.

Acknowledgements: Preparation of this paper was made possible in part by a Fellowship from the University of Southern California Center for Interdisciplinary Research. Travel to the Harvard meeting was funded by a grant from VPRO Dutch Television for travel to the Netherlands to tape a program on mirror neurons. Some of the material presented here was further developed for a pre-conference workshop entitled "Conversations on language, brain, culture" held in association with ASFLA 2002, Macquarie University, Sydney, Australia, on July 4th, 2002. I want to express my gratitude to the two referees and, especially, the editor for the probing comments that led me to recast the original draft in a form that greatly clarifies what I am trying to say.

References

Arbib, M. A., 1981, Perceptual structures and distributed motor control, in Handbook of Physiology – The Nervous System II. Motor Control (V. B. Brooks, Ed.), Bethesda, MD: American Physiological Society, pp.1449-1480.

Arbib, M.A., 2002, The Mirror System, Imitation, and the Evolution of Language, in Imitation in Animals and Artifacts, (Chrystopher Nehaniv and Kerstin Dautenhahn, Editors), The MIT Press, pp. 229 - 280.

Arbib, M.A., 2003a, Schema Theory, in The Handbook of Brain Theory and Neural Networks, (M.A. Arbib, Ed.), Second Edition, Cambridge, MA: A Bradford Book/The MIT Press, 993-998.

Arbib, M.A., and Bota, M., 2003, Language Evolution: Neural Homologies and Neuroinformatics, *Neural Networks (in press).*

Bickerton, D., 1995, Language and Human Behavior, Seattle: University of Washington Press.

Burgess, N., Jeffery, K.F., and O’Keefe, J. (Eds.), 1999, *The Hippocampal and Parietal Foundations of Spatial Cognition*, Oxford: Oxford University Press.

Byrne R. W., 2003, Imitation as behaviour parsing, *Phil. Trans. R. Soc. Lond. B*, 358:529–536.

Chomsky, N., 1965, Aspects of the Theory of Syntax, Cambridge, MA: The MIT Press.

Chomsky, N., 1992, A minimalist program for linguistic theory. In: Hale K, Keyser SJ (Eds.) The view from building 20: Essays in linguistics in honor of Sylvan Bromberger. The MIT Press, Cambridge MA , pp 1-52

Cobas, A., and Arbib, M., 1992, Prey-Catching and Predator-Avoidance in Frog and Toad: Defining the Schemas. *J. Theor. Biol*, 157:271-304.

Corballis, M. C., 2002, From hand to mouth: The origins of language. Princeton University Press.

Corballis, M.C., 2003, From mouth to hand: Gesture, speech, and the evolution of right-handedness, Behavioral and Brain Sciences (in press)

Darwin, C., 1872/1965, *The expression of the emotions in man and animals*. Chicago: University of Chicago Press.

Deacon, T.W.,1997, *The Symbolic Species: The co-evolution of language and the brain*, W.W. Norton & Company, New York & London.

Fagg A.H., Arbib M.A., 1998, Modeling parietal-premotor interactions in primate control of grasping. Neural Networks, 11: 1277-1303

Fogassi, L., Gallese, V., Fadiga, L. & Rizzolatti, G., 1998, Neurons responding to the sight of goal directed hand/arm actions in the parietal area PF (7b) of the macaque monkey. Soc. Neurosci. Abstr. 24, 257.

Ghazanfar, A.A., (Ed.), 2003, *Primate Audition: Ethology and Neurobiology*, Boca Raton: CRC Press.

Hauser M.D., Chomsky, N., Fitch, W.T. (2002) The faculty of language: what is it, who has it, and how did it evolve? Science. 298:1569-79.

Hewes, G., 1973, Primate communication and the gestural origin of language. Current Anthropology, 14:5-24.

Hurford, J.R., 2003, Language beyond our grasp: what mirror neurons can, and cannot, do for language evolution, in Evolution of Communication Systems: A Comparative Approach, (D. Kimbrough Oller and Ulrike Griebel, Eds.), Cambridge, MA: The MIT Press.

Jackendoff , R., 2002, Foundations of Language: Brain, Meaning, Grammar, Evolution, Oxford and New York: Oxford University Press,

Kohler E, Keysers C, Umilta MA, Fogassi L, Gallese V, Rizzolatti G., 2002, Hearing sounds, understanding actions: action representation in mirror neurons, Science, 297(5582):846-8

Lieberman, P., 2000, Human Language and Our Reptilian Brain, The Subcortical Bases of Speech Syntax and Thought, Cambridge, MA: Harvard University Press.

MacNeilage, P.F., 1998, The frame/content theory of evolution of speech production. Behav Brain Sci 21:499-546.

Oztop, E., and Arbib, M.A., 2002, Schema Design and Implementation of the Grasp-Related Mirror Neuron System, Biological Cybernetics, 87:116–140.

Oztop, E., Bradley, N., and Arbib, M.A., 2002, Learning to Grasp I: The Infant Learning to Grasp Model (ILGM) (to appear).

Passingham, R., 1993, *The Frontal Lobes and Voluntary Action,* Oxford: Oxford University Press.

Rizzolatti G, Fadiga L, Gallese V, Fogassi L., 1995, Premotor cortex and the recognition of motor actions. Cogn Brain Res, 3: 131-141

Rizzolatti G., Arbib M.A., 1998, Language Within Our Grasp. Trends in Neurosciences, 21(5): 188-194

Stokoe W. C., 2001, Language in Hand: Why Sign Came Before Speech, Washington, DC: Gallaudet University Press.

Supalla, T., and Newport, E., 1978. How Many Seats in a Chair? The Derivation of Nouns and Verbs in American Sign Language. In Understanding Language through Sign Language Research, (P. Siple, Ed.) New York: Academic Press, pp.91-159..

Taira M, Mine S, Georgopoulos AP, Murata A, Sakata H (1990) Parietal Cortex Neurons of the Monkey Related to the Visual Guidance of Hand Movement. Experimental Brain Research, 83: 29-36.

Umilta MA, Kohler E, Gallese V, Fogassi L, Fadiga L, Keysers C, Rizzolatti G., 2001, I know what you are doing. a neurophysiological study, Neuron 31(1):155-65.

Zukow-Goldring, P., Arbib, M.A., and Oztop, E., 2002, Language and the Mirror System: A Perception/Action Based Approach to Communicative Development (to appear).